Article

A Segmentation Model for Extracting Farmland and Woodland from Remote Sensing Image

Chengming Zhang [1,2,*], Shujing Wan [3], Shuai Gao [4], Fan Yu [2], Qingdi Wei [1,5], Guang Wang [1], Qing Cheng [1,5], Dejuan Song [1,5]

[1] College of Information Science and Engineering, Shandong Agricultural University,61 Daizong Rd,271000, Taian, Shandong ,China; chming@sdau.edu.cn; 904840865@qq.com
[2] Chinese Academy of Surveying and Mapping, 28 Lianhuachixi Rd,100068, Beijing ,China; yufan@casm.ac.cn
[3] Network Information center, QuFu Normal University, 57 Jingxuan Rd, 273165, Qufu, Shandong, China; wanshujing89@163.com
[4] Institute of Remote Sensing and Digital Earth, Chinese Academy of Sciences; gaoshuai@radi.ac.cn
[5] Shandong Technology and Engineering Center for Digital Agriculture, 61 Daizong Rd,271000, Taian, Shandong ,China; 1064190252@qq.com; 1044241223@qq.com; 2416871293@qq.com
* Correspondence: sdauzcm@163.com; Tel.: +86-139-5382-3659

Abstract: It is very difficult to accurately divide farmland and woodland in Gaofen 2 (GF-2) remote sensing image, because their single plant coverage is very small, and their spectra are very similar. The ratio of spatial resolution and one plant's coverage area must be fully taken into account when designing the Convolutional Neural Network structure for extracting them from GF-2 image. We establish a Convolutional Encode Neural Networks model (CENN), The first layer has two sets of convolution kernels to learn the characteristics of farmland and woodland respectively, while the second layer is the encoder to encode the characteristics by transfer function, which can map the results to the corresponding category number. In the training stage, samples of farmland, woodland, and other categories are categorically used to train CENN, as soon as training is accomplished, CENN would acquire enough ability to accurately extract farmland and woodland from remote sensing images. The final extraction result is obtained by implementing per-pixel segmentation of images used to train the CENN. CENN is compared and analyzed with others such as Deep Belief Network (DBN), Full Convolutional Network (FCN), Deeplab Model. The results of experiments show that CENN can more accurately mine the characteristics of farmland and woodland, and it achieves its goal of extracting farmland and woodland with high precision from GF-2 images.

Keywords: Convolutional Neural Network; Gaofen 2 remote sensing image; remote sensing image segmentation; Convolutional Encode Neural Networks model (CENN); categorical learning; per-pixel segmentation; farmland; woodland

1. Introduction

Image segmentation is the precondition and foundation for the extraction and target identification of high-resolution remote sensing image [1]. In high-resolution images, the spectral features are more abundant, In the high-resolution image, The spectral confusion is more serious and the differentiation is significantly reduced, and the accuracy of the segmentation method based on spectral statistics is significantly reduced[2,3].Object-oriented image segmentation method can overcome the influence of "salt and pepper" noise and improve the accuracy by using object structure and spectral signature. It needs to adjust the segmentation scale to obtain the suitable image segmentation result, but the suitable segmentation scale is difficult to be determined, which resulted in the slow development of the object-oriented segmentation method [4, 5].

With the development of machine learning technology, researchers began to apply machine learning algorithms such as Neural Networks (NN) [6,7] and Support Vector Machine (SVM) [8,9] to

 doi:10.20944/preprints201712.0192.v1

the segmentation of high-resolution images and added Image texture, structure, and other features to improve the segmentation accuracy [10-12]. The study shows that image segmentation based on machine learning algorithm can obtain more optimal results compared with the traditional statistical methods and object-oriented methods [13,14]. But both SVM and NN belong to the shallow learning algorithm [15-17], Due to the limited network structure, shallow learning algorithm has difficult to express complex function effectively, so enhanced with the increase of sample size and diversity of samples, shallow model also gradually cannot adapt to the complex samples.[18,19].

The recent advancement in the deep learning motivates us to address these problems with deep neural networks [20–23]. As one of the important branches of deep learning, convolutional neural network has long been applied to image data due to its outstanding ability of feature learning [24–26].The deep learning network composed of multi nonlinear mapping layers has strong function expression power which has obtained excellent results in image segmentation and many of its achievements have been approved [27,28]. As one of the important branches of deep learning, Convolutional Neural Network has long been applied to image data due to its prominent ability in feature learning. The traditional deep learning methods involved deep convolutional neural networks (DCNN) [29,30], deep deconvolutional neural networks (DeCNN) [31].Since then, many methods of remote sensing image segmentation based on Convolutional Neural Network have been developed[32,33]. Many large convolutional neural network whose performance can be scaled depending on the size of training data, model complexity as well as processing power, has shown significant improvements in object segmentation from images [34-41].

Fully Convolutional Networks (FCN) is a deep learning network for image segmentation proposed in 2015 [41]. Taking benefit of the advantages of convolutional computation in feature organization and extraction, the network establishes a multi-layer convolution structure and reasonable sets deconvolution layer to realize explicit segmentation [42-44]. Since then, researchers have developed a series of segmentation models based on convolution, such as segNET [45], UNet [46], DeepLabel [47], Multi-Scale FCN [48], reSeg [49], which have their own strengths and work very well in different types of images.

The key models such as FCN achieve success because they express rich detail features on images well through the multi-layer convolution. But we notice that in farmland region, one pixel usually contains several individual plants, thus the information between pixels does not change much and the image texture is smoother. Although the individual trees are larger than the crop, a tree typically occupies two or three pixels on the GF-2 image, and smaller trees tend to occupy only one pixel or less than one pixel. The information change between pixels is still small, due to the difference of the pixel content, the image has a grainy texture. Because of the nature of the pixel that covers the crops and trees, the effect of deep layer convolution is very small, and more features cannot be found, and even greater noise can be pulled in, resulting in poor segmentation effect.

From the above analysis, we may find that the proportion of spatial resolution of the image and the area size of extracting features must be fully considered when image segmentation by deep learning is implemented in order to design a network structure for the specific scenario, to improve the segmentation accuracy. Overall, the paper puts forward a Convolutional Encode Neural Networks (CENN) model according to the characteristics of the GF-2 images, such as lesser covering area of single crop and tree, fewer pixels and details and continuous plant emergence to realize farmland and woodland extract, which improve the segmentation accuracy effectively and achieves good results. The paper mainly carried on the following work:

(1) According to the characteristics of GF-2 data, a network structure composed of convolution and encoder is designed, in which the convolution part is used to extract the features, the encoder is used to encode and map them into the corresponding feature categories.

（2）In the model training, the method of classification training was adopted. Taking the sample of farmland and woodland as the positive sample, the others as the negative sample, and the farmland and the woodland as negative samples for each other, the model was trained to obtain sufficient recognition ability.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 28 December 2017 doi:10.20944/preprints201712.0192.v1

（3）Use the trained model to segment the image and evaluate the accuracy of the segmentation results.

2. Methods

Similar to other research on image segmentation using convolutional neural networks, we divide the work into two parts of training stage and classification stage by using CENN, as shown in Figure 1.

The upper part of Figure 1 shows the training stage, image-label pairs, with pixel-class correspondence, which together becomes input to the CENN as training samples. The error between predicted class labels and labels is calculated and back-propagated through the network using the chain rule, and then the parameters of the CENN are updated using the gradient descent method. The above iteration will be stopped when the error is less than a given threshold. The lower part of Figure 1 shows the classification stage, the trained CENN has enough ability exacting farmland and woodland accurately on an input image.

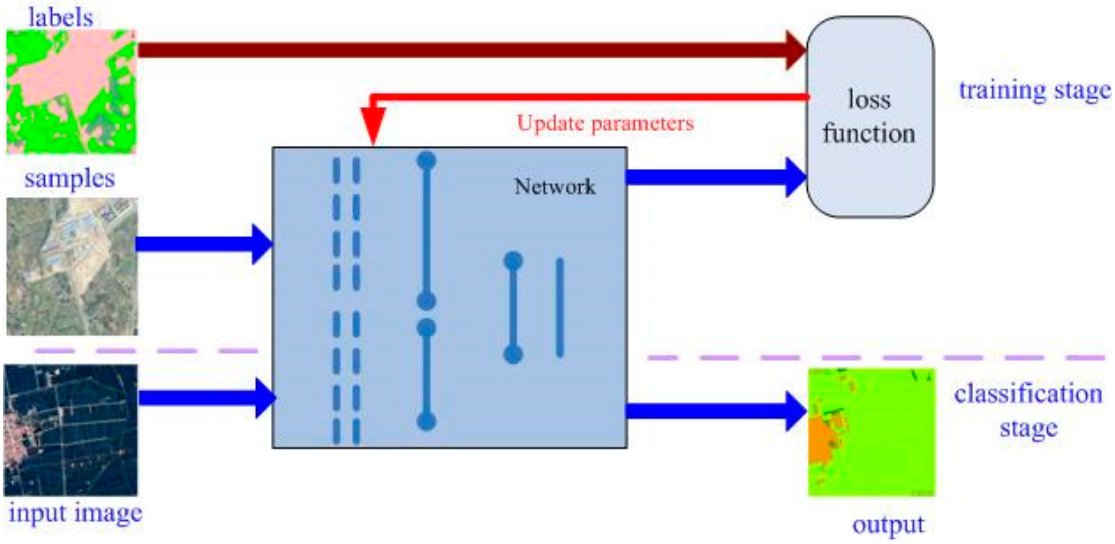


Figure 1. The frames of our research

2.1 Network Architecture

CENN model is divided into four parts as Figure 2 shows, (1) input; (2) convolution kernel group; (3) encoder group; (4) output. In the training stage, the input is original images and labels; In the classification stage, the input is the original image; The output is a single-band file, and the content of each pixel is the class number of the corresponding original image. In this model, the category number 100 is farmland, 150 is woodland and 200 is uncategorized. Let's focus on convolution kernel and encoder group.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 28 December 2017 doi:10.20944/preprints201712.0192.v1

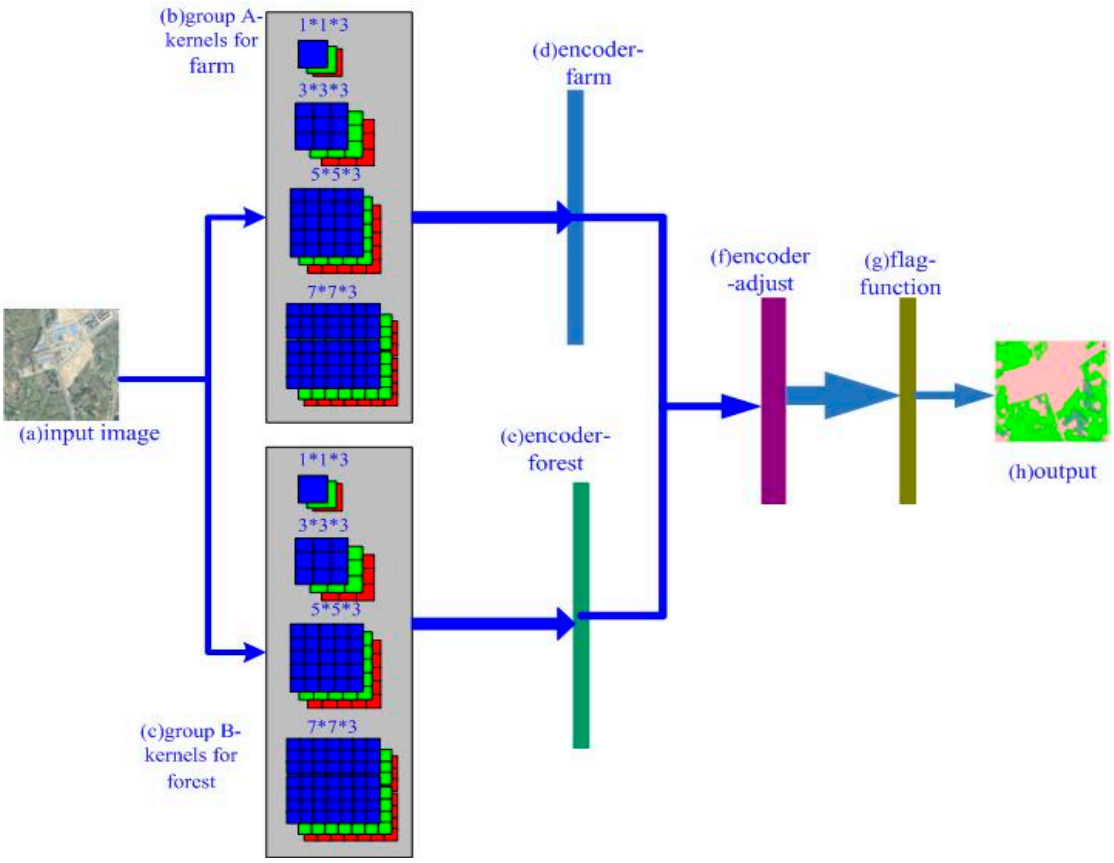


Figure 2. The network architecture of CENN

2.1.1 Convolution Kernel Group

As in the previous analysis, the effect of deep convolution is very small because of the characteristics of farmland and woodland data. The convolution part of CENN model does not use deep convolution but uses the "width" convolution model to extract more local features to enhance the image pixel segmentation accuracy of the image pixel. The convolution kernels of the CENN model are in r*c*h form, where r denotes the width of the convolution kernel, c denotes the height of the convolution kernel, and h denotes the number of channels of the convolution kernel. In this paper, h are all set to 3 because we only use three channels of GF-2. There are four types of convolution kernel which is 1*1*3，3*3*3，5*5*3，7*7*3, designed in CENN as Figure 2（b）(c) shows.

(1) 1*1*3The main role of the convolution kernel is to extract the color characteristics of the center pixel.

(2)The convolution kernel of 3*3*3 is divided into two sub-categories. The first subclass is the convolution kernel that needs to be trained, and this type of convolution kernel is used to extract the texture features of the central pixels and the surrounding 8 pixels. The second subclass is a fixed convolution kernel, which is used to extract the color change between two adjacent pixels. It is further divided into eight sections, as shown in Figure. 3. The eight convolution kernels are used to calculate the value of the color difference between the central pixel and the upper left, upper, upper right, left, lower left, lower right and lower right eight adjacent pixels. The absolute value symbol|-1|in it indicates that the calculation result is the absolute value of the color difference, so it will be convenient for subsequent operations.

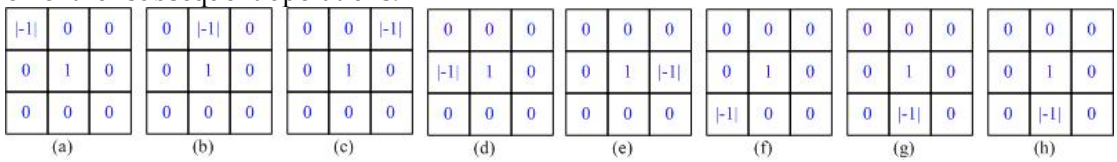


Figure 3. Eight fixed convolution kernel of 3*3*3

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 28 December 2017 doi:10.20944/preprints201712.0192.v1

(3) The function of 5*5*3, 7*7*3 types of convolution kernels is similarto the 3*3*3 types of convolution kernels, but they have a wider range to exploit the feature between center pixels and surrounding pixels. Similar to the convolution kernels of 3*3*3 types, convolution kernels of 5*5*3 and 7*7*3 each have two subclasses. Figure 4 shows 8 fixed convolution kernels of 5*5*3 and Figure 5 shows 8 fixed convolution kernels of 7*7*3.

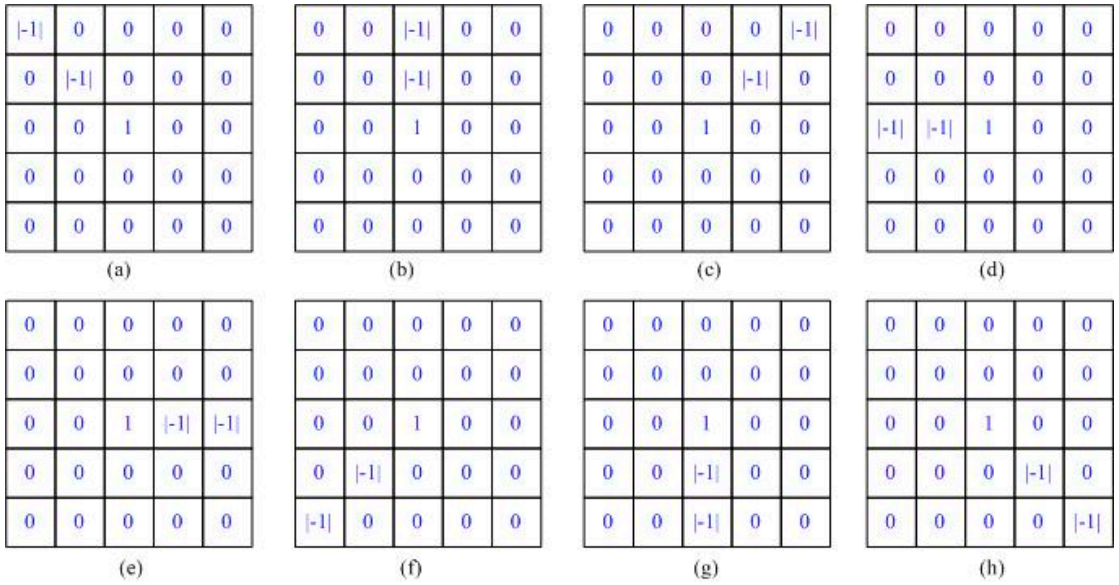


Figure 4. Eight fixed convolution kernel of 5*5*3

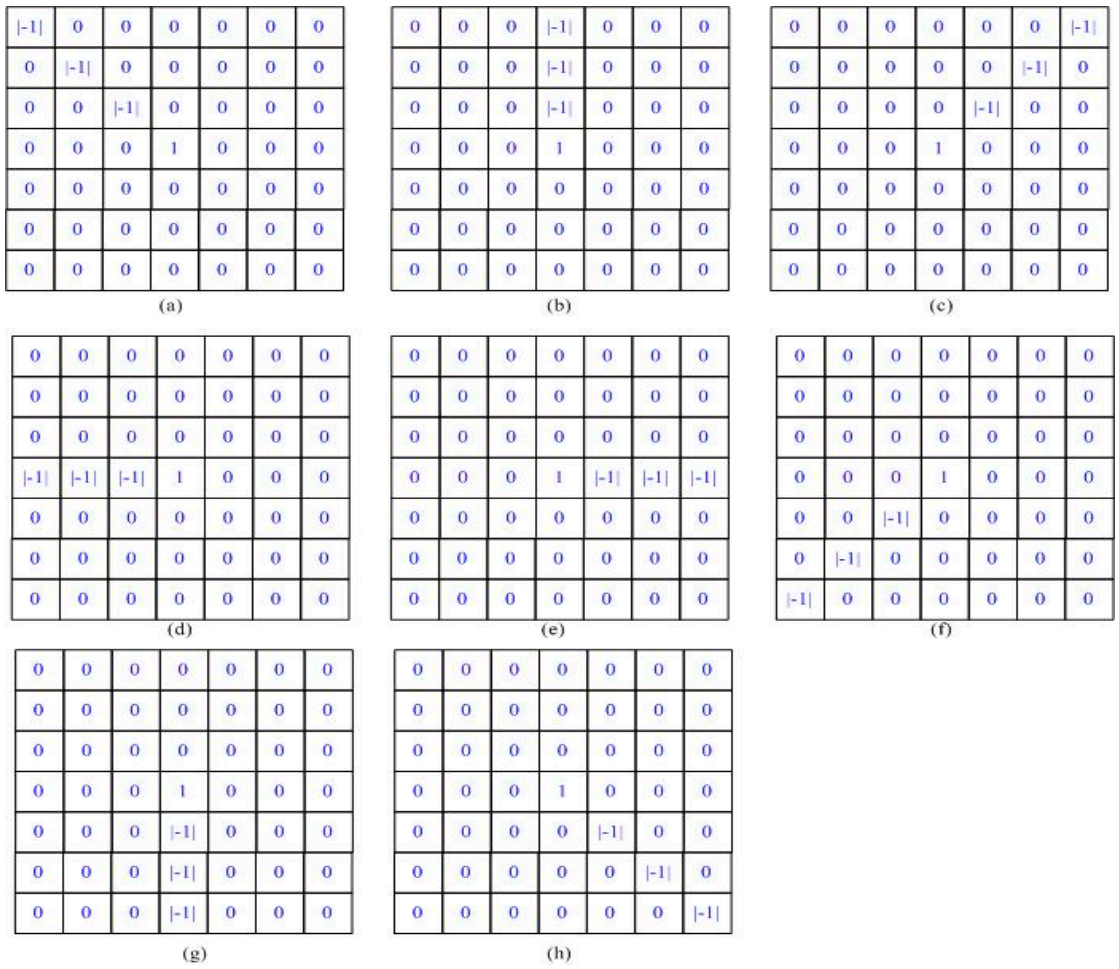


Figure 5. Eight fixed convolution kernel of 7*7*3

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 28 December 2017 doi:10.20944/preprints201712.0192.v1

In Fig. 4 and Fig. 5, the convolution result shows the sum of the differences between the pixels in each direction, and the convolution result is averaged to correspond to the result of the fixed convolution kernel of 3*3*3.

Among them, 5 * 5 * 3 processing method is:

$$r = \frac{r_1 + r_2}{2} \quad (1)$$

7 * 7 * 3 is handled as follows:

$$r = \frac{r_1 + r_2 + r_3}{3} \quad (2)$$

r_1, r_2, r_3 means differences between the pixels.

The convolution kernels were divided into two groups A and B, group A was designated to learn farmland samples, and group B to study woodland samples. Both groups contain 1*1*1，3*3*3，5*5*3，7*7*3 four type convolution kernels. After training, two sets of convolution kernels are obtained. The purpose of this design is that the model can better express the unique characteristics of farmland and woodland, and help to improve the distinction ability.

2.1.2 Encoder sets

The encoder set is designed with two layers of encoders to fit the non-linear relationship between extracted features and output targets.

There are two encoders in the first encoder layer, that is the (d)encoder-farm and the (e)encoder-forest in Figure. 2. The (d)encoder-farm is used for performing regression fitting on the characteristics of the farmland obtained from learning, and the (e)encoder-forest is used for fitting the characteristics of woodland obtained from learning.

There is 1 encoder in the second encoder layer, that is the (f)encoder-adjust in Figure. 2. The function of the (f)encoder-adjust is to adjust the (d)encoder-farm and the (e)encoder-forest, make the fitting results of the two interval for complete separation, and to ensure that the negative sample information after the calculation of (d)encoder-farm or (e)encoder-forest, does not appear on the farmland and woodland interval. Then the (g)flag-function maps the encoded result of (f)encoder-adjust to the category number, as shown in Figure 6.

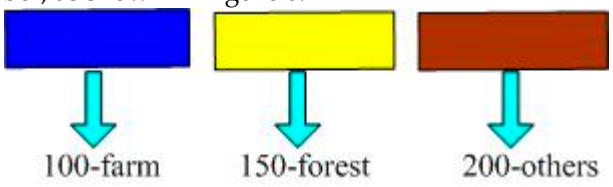


Figure 6. Result schematic diagram of encoder 3

In Figure. 6, the blue part represents the farmland coding result, the yellow part represents the woodland coding result, and the brown part represents the negative sample coding result.

2.2 Network Training

2.2.1 Label sample

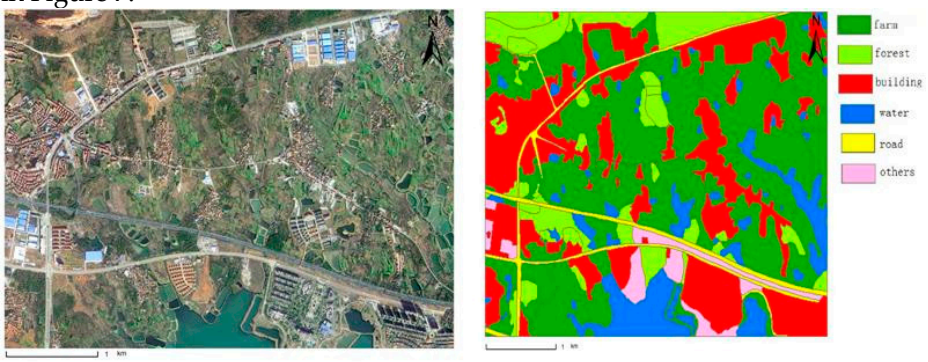
In our training dataset, there are a total of 28 remote sensing images of GF-2(size 7300*6900) of Shandong Province, including 12 images of February 17, 2016, and 16 images of May 12, 2017. The panchromatic band of the spatial resolution is 0.8 meters, the multi-spectral spatial resolution is 3.2 meters. We used ENVI software to do some pretreat work such as fusion, color stretching for the panchromatic band and 4 multi-spectral bands, and selected 321 as RGB bands, to improve visual effects.

Artificial label samples are very important foundations for training. Since the CENN model uses pixels as the main learning object, it must be accurate when labeled. We use ENVI software for labeling and implement preprocessor to make the mask. The process of the artificial label is:(1) In the ENVI software, the ROI (Region of Interest) tool is used to select farmland and woodland in the image data in turn, and select other regions as the un-classified regions. The results, only contained Map

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 28 December 2017 doi:10.20944/preprints201712.0192.v1

Location of the pixel, named image coordinates, were output to three regions of interest documents. (2)The steps of preprocessor are as follows: adding a new band as mask band in the image file, and the spatial resolution and size of the new band are the same as the original image; reading ROI files in turn, write the class number to the corresponding position in the mask band according to the Map Location of each pixel.

We manually labeled all images at the pixel level as label data. In other words, for each image, there exists a 7300*6900 label map, having a pixel-class (row-col indexed) correspondence with it. We used 20 images for training and the remaining 8 images for testing. One label pair examples are illustrated in Figure 7.



(a)image (b)labels

Figure 7. Image-label pair example (a) Original image (b) labels

2.2.2 Model Training

We chose two periods of the image as training data. The reason why we chose images with a different period is to increase the anti-interference abilities of our model, such as the change of seasons, to enhance its applicability. The general procedure of our training stage is: image-label pairs are input into the CENN as training samples. The Encode function is performed on the output feature map generated by the network to predict the class distribution. Then the cross-entropy loss is calculated and back-propagated, and finally, the network parameters are updated using Stochastic Gradient Descent (SGD) [43, 50] with momentum. S

After the training is completed, two sets of Convolution Kernels A and B will be obtained. In Group A, farmland features may be enhanced while other types of features are suppressed as much as possible. In Group B, woodland features may be enhanced while other types of features may be suppressed.

In our training, the SGD method with momentum is used for parameter updates, the following expression illustrates the SGD [43, 50] method with momentum.

$$W^{(n+1)} = W^{(n)} - \Delta W^{(n+1)} \quad (3)$$

Where, W(n) denote the old parameters while W(n+1) denotes the new parameters, and $\Delta W(n+1)$ is the increment for the current iteration, which is a combination of old parameters, gradient, and historical increment:

$$\Delta W^{(n+1)} = \vartheta\left(d_w \cdot W^{(n)} + \frac{\partial J(W)}{\partial W^{(n)}}\right) + m \cdot \Delta W^{(n)} \quad (4)$$

Where, $J(W)$ is the loss function, ϑ is the learning rate for step length control, d_w denote the weight decay, and m denote the momentum.

2.4 Segment using the trained network

After training, the model can be used to segment the input image pixel by pixel. According to our design, the output will be written into a new band with the value of each element in this band as the category number of the corresponding original pixel. As mentioned earlier, three class numbers were used in our experiments: 100 - Farmland, 150 - Woodland, 200 - Uncategorized. The benefit of this design is that it saves the segmentation result and avoids any damage to the original file.

 doi:10.20944/preprints201712.0192.v1

3. Experiments

We designed a set of test experiments and comparative experiments to verify the feasibility of CENN. Our algorithm is implemented using Python 2.7 and is performed on the Linux Ubuntu 16.04 operating system installed NVIDIA GeForce Titan X Graphics device with 12G byte graphic memory.

The data and classification criteria we used have already been described in Section 2.2.1. We selected about a half pixels of each image as samples to train CENN and the rest as test data.

3.1 Learning ability of CENN indicators

The ability of CENN model is mainly reflected in the two aspects of feature extraction ability and coding ability. The features of the four scales extracted by CENN model are all in numerical form. It accumulates the degree of concentration of the extracted eigenvalues as an index to examine the feature extraction ability, and takes the discrimination among farmland, woodland and background as an index to check the coding ability after coding, further the proportion of the number of the wrong pixel as the index of a test model is taken to examine the overall segmentation ability of the model.

3.2 Comparison Model

We chose the DBN model, the FCN model and the Deeplab model as the comparative models and used specific methods given in the open literature for comparative experiments.

3.2.1 DBN

This paper [2] gives a method of pixel-by-pixel classification for high-resolution images by using DBN. The method uses non-subsampled contourlet transform to calculate the texture features of the image and uses DBN to classify the high-resolution remote sensing images based on spectral-texture features.

The training process of this thesis includes two processes: pre-training and fine-tuning. The pre-training is conducted in an unsupervised manner through unlabeled samples. During training, the greedy algorithm is used to optimize layer by layer. The parameters of Restricted Boltzmann Machines(RBM)in each layer are adjusted independently. After training one layer, the output of the layer is regarded as the input of the next layer, and the training of the next RBM is continued. After the pre-training, a supervised learning mode is used to train the last layer of network, and the error is propagated layer by layer. The weight of the entire DBN network is fine-tuned, and the backpropagation method is used during this process.

3.2.2 FCN

For the FCN model, we directly employ the FCN-8s model proposed by Jonathan Long et al. [42].The architecture of the model is also the VGG-16 network. The final prediction is fused from the output of three branches (from the primary network, the pool4 layer, and the pool3 layer, respectively) after the up-sampling operation. In the training phase, the input data and the training parameters for FCN-8s in the experiment are the same as ours. In the testing stage, we use the same classification parameters as our approach.

3.2.3 DeepLab

For the DeepLab model, we directly employ the DeepLabel v3 model proposed by Liang-Chieh Chen et al. [47]. DeepLab is also based on VGG network like FCN, but there is a difference. In order to ensure that the output size will not be too small and without adding too much padding, Deeplab used a very elegant approach: the stride of pool4 and Pool 5 layers in VGG network changed from the original 2 to 1, plus 1 padding. This modification makes the total stride of VGG network change from 32 to 8, which makes fc7 get the score map of 67x67 when the input image is 514x514. It is much denser than FCN. To compensate for the effects of stride changes on the receptive field, Deeplab uses a convolution method called "Atrous Convolution" to ensure that the pooled experience remains

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 28 December 2017 doi:10.20944/preprints201712.0192.v1

unchanged and the output is more refined. Finally, DeepLab takes a Fully-Connected Conditional Random Fields to refine the split boundary.

3.3 Results and Comparison

3.3.1 Learning ability of CENN

In Figure 8, we give the distributions of the features learned from Sample-Farm, Sample-Forest, and Sample-Ground respectively. It can be seen from Figure 9 that after the convolution operation, the characteristics of farmland and forest are mainly concentrated in two regions while the Sample-Ground type is very scattered. This is mainly because of the seasonal differences between the data we selected and the more significant differences in color values. At the same time, the characteristics concentration of the forest is less than that of farmland, which is mainly because the color change of woodland is much larger than that of farmland.

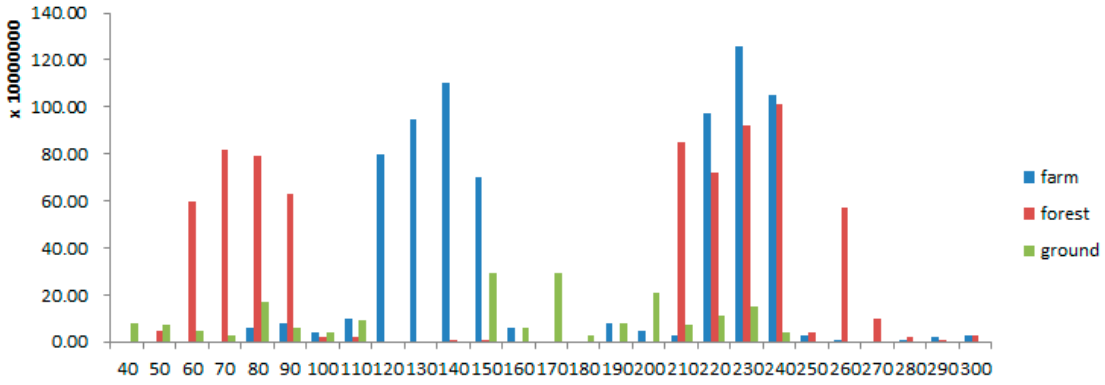


Figure 8. Training results of 1*1*3 type convolution kernels

Figure 9 also shows the learning results of 3*3*3 convolution kernels for Sample-Farm, Sample-Forest, Sample-Ground in the form of a histogram. As it is evident from the figure, since the 3*3*3 convolution kernel is mainly used to learn the color difference between adjacent pixels, the features of Sample-Farm samples have better concentration degree, and the dispersion level of Sample-Forest are significantly larger, reflecting the features that the texture of Sample-Farm are fine and the texture of Sample-Forest are rough.

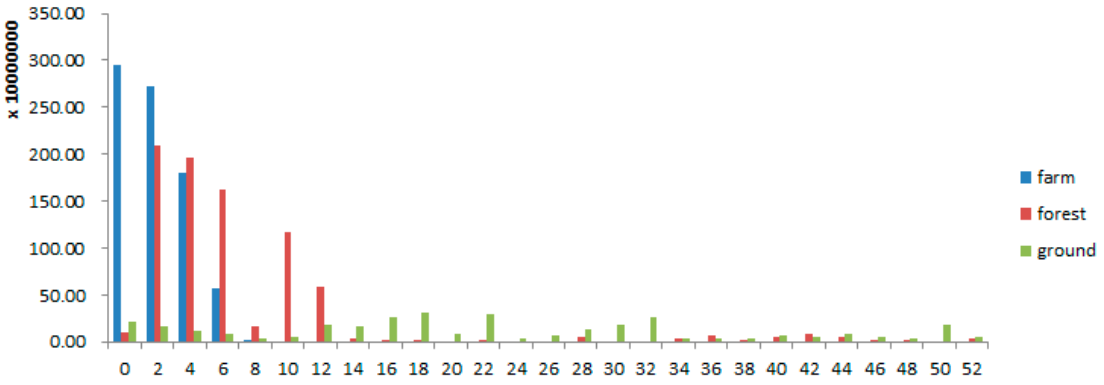


Figure 9. Training results of 3*3*3 type convolution kernels

Further, for farmland, the characteristics concentration of 5*5*3 convolution kernels and the characteristics concentration 7*7*3 convolution kernels both less than the characteristics concentration 3*3*3 convolution kernels, The reason is that as the field of vision expands, more negative sample pixels are introduced into the boundary. But Sample-Forest5*5*5 has the best

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 28 December 2017 doi:10.20944/preprints201712.0192.v1

concentration of features, this shows that the convolution kernel of 5*5*3 is more suitable for the extraction of woodland features. Figure 10 shows the learning results of 5*5*3 convolution kernels.

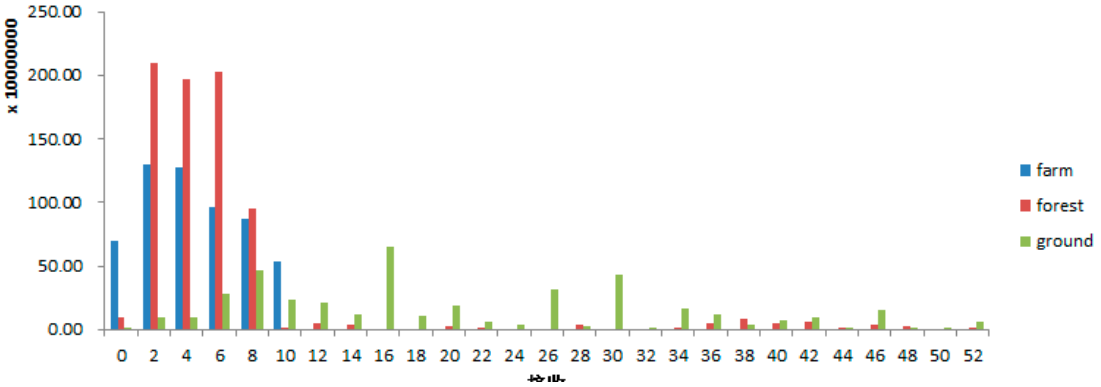


Figure 10. Training results of 5*5*5 type convolution kernels

From the results, we can observe that on one hand, multi-space convolution kernel is more specifically suitable for the extraction of farmland and woodland features rather than the application of depth convolution kernel, on the other hand, the result indicates the need to combine multiple features in order to accurately determine the type of pixel description, and the necessity of encoder.

Figure 11(a) shows the encoding result of the first layer encoder, (b) shows the encoding result adjusted by the second layer encoder. It can be seen from the figure that the adjusted encoding result is already available to use for Segmentation of the pixels.

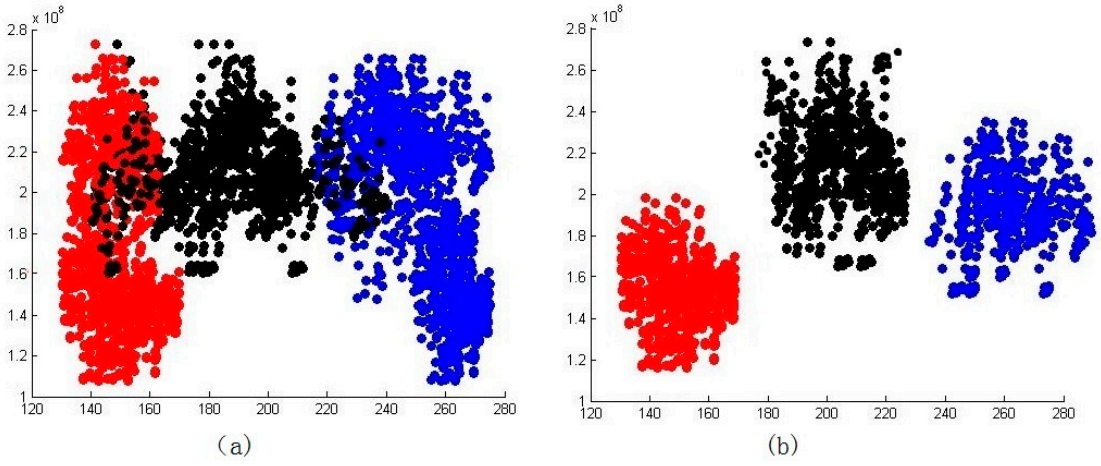


Figure 11. Encoding result

3.3.2 Experiment Result Comparison

In Comparison Experiment, we adopt our trained model on four GF-2 for the Segmentation. All the image sizes are 7300*6900. These images are the testing images that are not involved in training. Figure 12 is the illustration of the results and the comparisons. Figure 13 is the Errors of DBN, Figure 14 is the Errors of FCN, Figure 15 is the Errors of Deeplab.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 28 December 2017 doi:10.20944/preprints201712.0192.v1

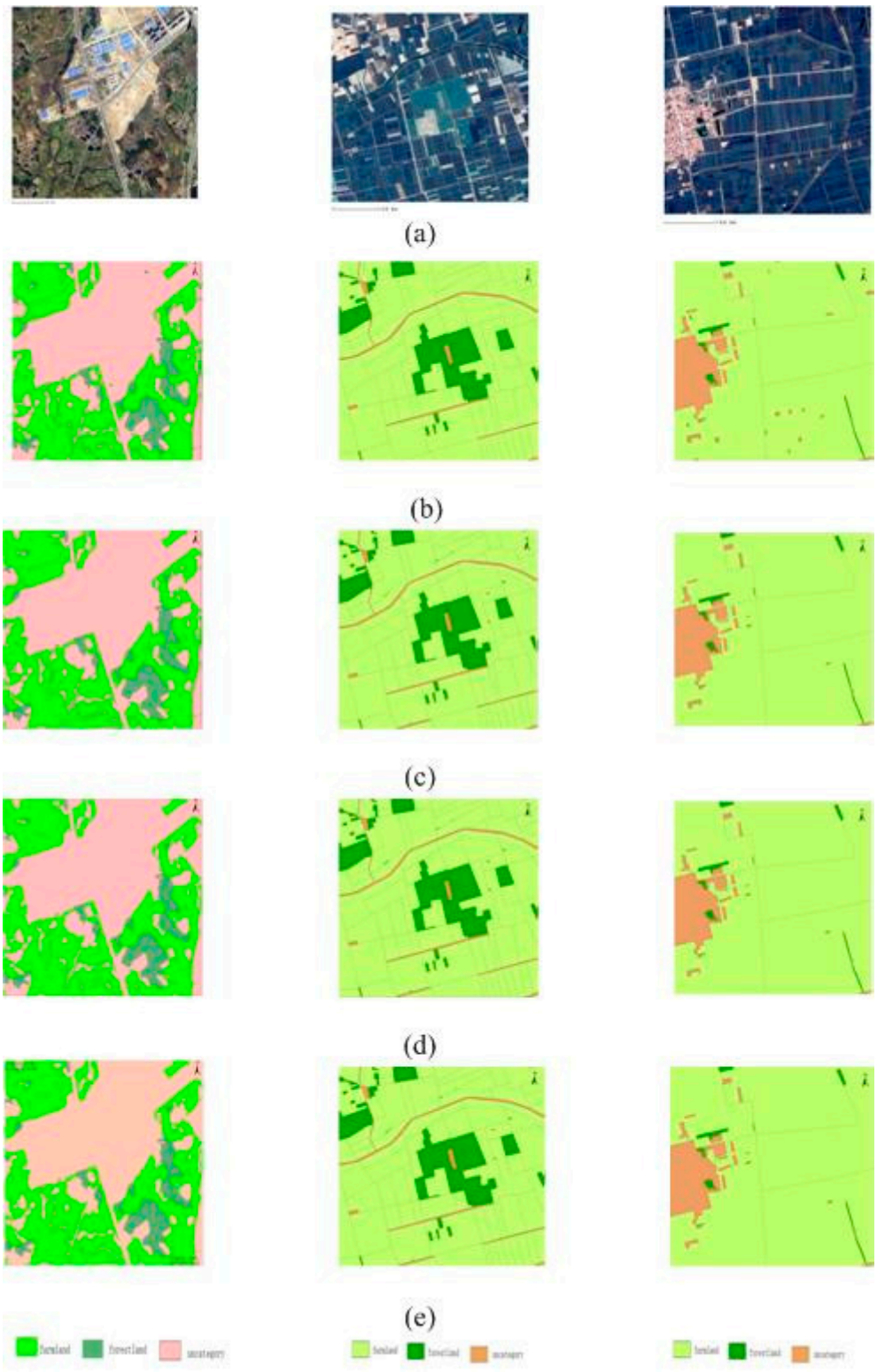


Figure 12. Segmentation results on GF-2 images. (a) Original images; (b) Results of DBN; (c) Result of FCN; (d) Results of Deeplab; (e) Our results corresponding to the images in (a).

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 28 December 2017 doi:10.20944/preprints201712.0192.v1

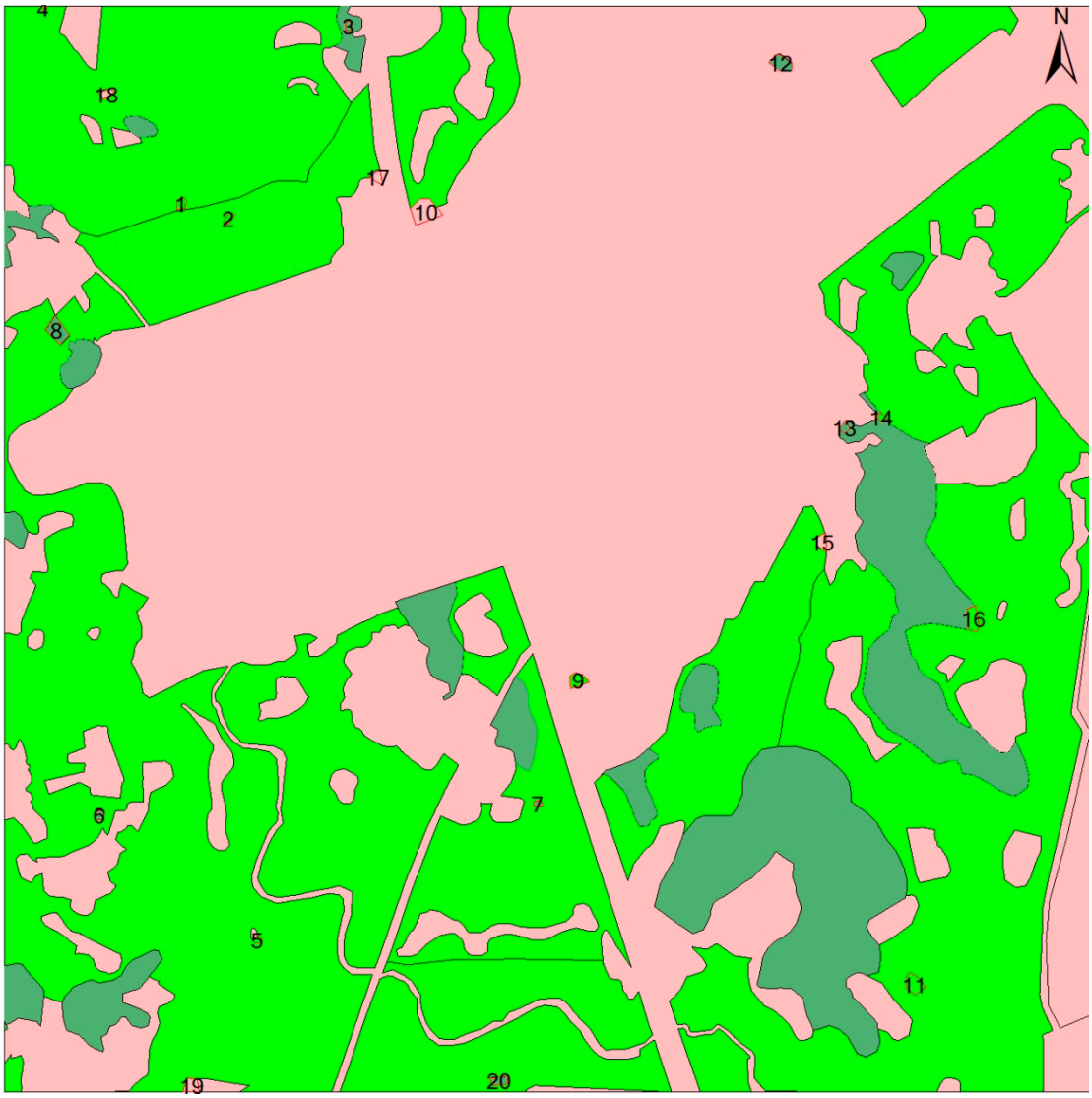
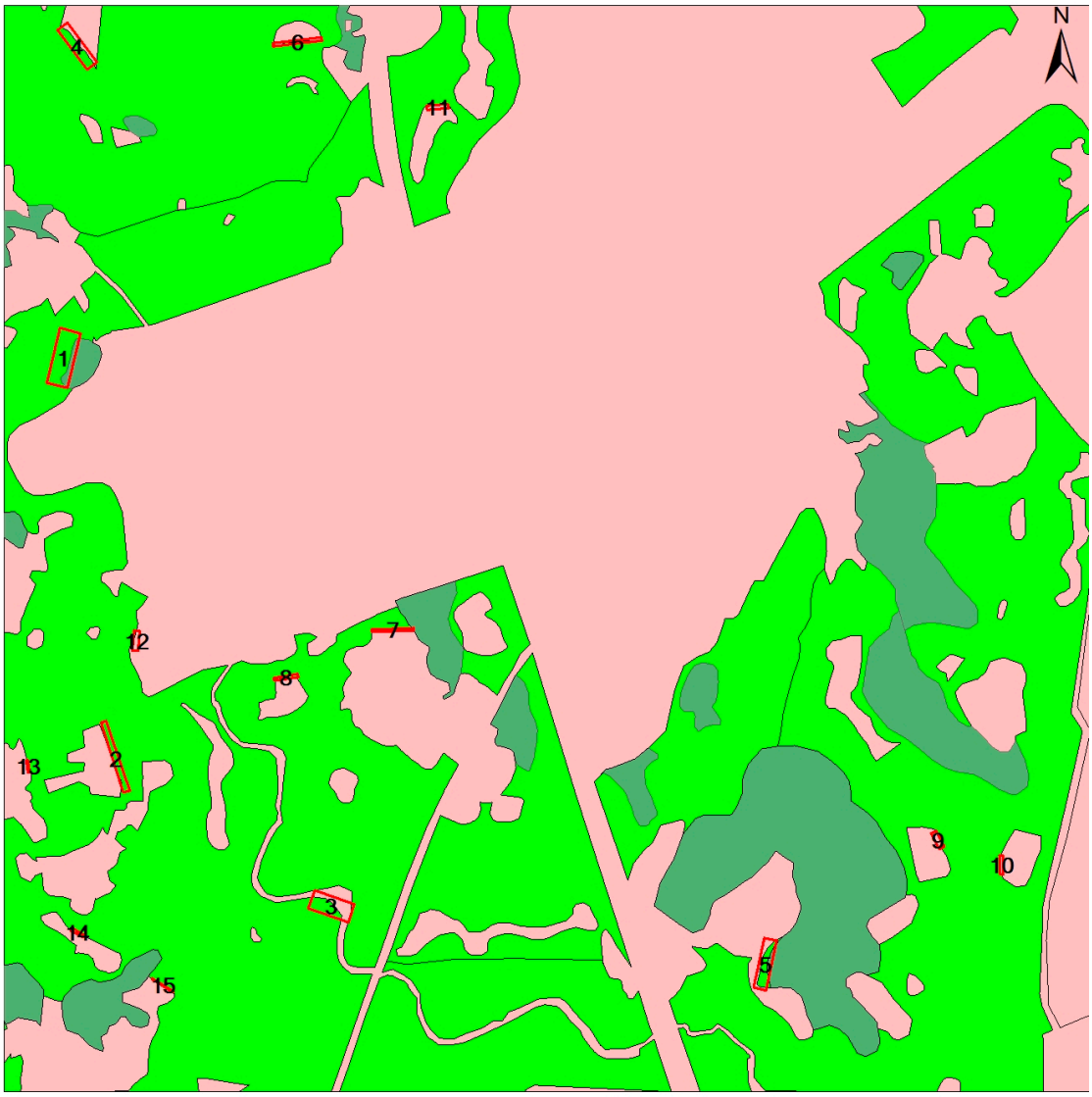


Figure 13. The Errors of DBN

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 28 December 2017 doi:10.20944/preprints201712.0192.v1



330

331 **Figure 14.** The Errors of FCN

332

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 28 December 2017 doi:10.20944/preprints201712.0192.v1

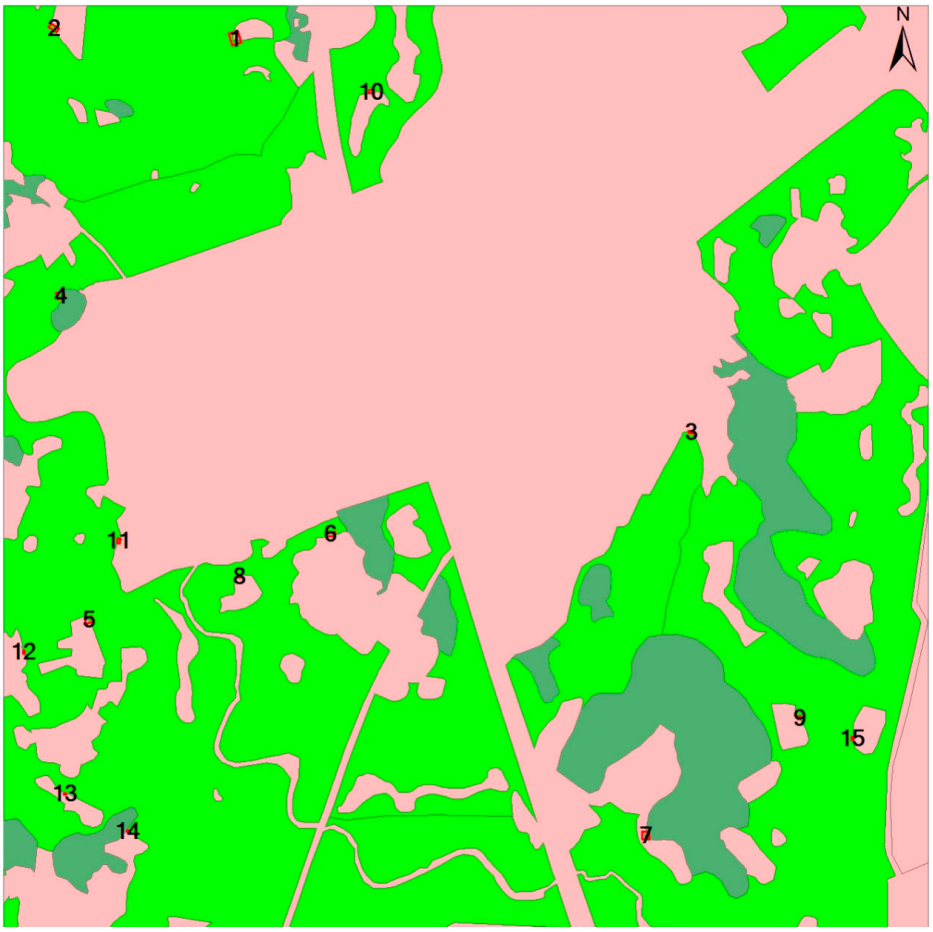


Figure 15. The Errors of Deeplab

We employ precision, recall, and Kappa coefficient as the indicators to evaluate our approach. These indexes are calculated from the confusion matrix C, where the precision is calculated as $\frac{1}{3}\sum_i C_{ii}/\sum_j C_{ij}$ that denotes the average proportion of pixels being classified to one class that are correct, and the recall is computed as $\frac{1}{3}\sum_i C_{ii}/\sum_i C_{ij}$ that represents the average proportion of pixels that are correctly classified, and the Kappa coefficient measures the consistency of the predicted classes with the manual labels [43]. The comparisons are listed in Table 1.

Table 1. Comparison between approaches using DBN, FCN, Deeplab, and CENN

Approach	Index	Experiment-1	Experiment-2	Experiment-3
DBN	Precision	0.69	0.74	0.76
	Recall	0.61	0.63	0.62
	Kappa	0.58	0.69	0.71
FCN	Precision	0.79	0.81	0.76
	Recall	0.72	0.75	0.69
	Kappa	0.71	0.73	0.64
Deeplab	Precision	0.84	0.86	0.79
	Recall	0.77	0.79	0.73
	Kappa	0.79	0.81	0.77
CENN	Precision	0.91	0.93	0.93
	Recall	0.85	0.88	0.87
	Kappa	0.82	0.85	0.84

 doi:10.20944/preprints201712.0192.v1

4. Discussion

This paper presents a classification approach，which extracts farmland and woodland from high-resolution images using the CENN model. Compared with the three typical deep learning-based approaches, the classification accuracy is obviously improved. In the following sections, we will discuss the reasons.

4.1 DBN vs. Our Approach

In the method of pixel segmentation based on DBN, the texture feature of image is computed first and the obtained two-dimensional texture features are transformed into a one-dimensional vector. Then the three channel values of RGB are added to this Vector and merged into one vector. Finally, each component value of the vector is taken as an independent input to construct a DBN network to achieve the purpose of classifying the pixels. The use of the texture features of the components disassembled did not play the role of the spatial relationship represented by the texture, thereby losing the meaning of texture extraction. Therefore, similar to traditional spectral-based methods, this method utilizes only the spectral characteristics of the pixels themselves and fails to effectively utilize the spatial relationship between pixels, resulting in easy judgment errors in applications.

Unlike the DBN method, the CENN model makes full use of the advantages of convolution in information aggregation and uses 1*1*3 convolution kernel to extract the common features of the original spectral value. Three kinds of convolution kernels of 3 * 3 *, 5 * 5 * 3 and 7 * 7 * 3 are used to extract textural variation characteristics of pixels in three sizes. Then, coding the four kinds of the same amount twice is logically clear, complete and accurate. Because CENN fully excavated many features of data, meanwhile, using the two-stage encoder to simulate nonlinear equations to encode the features, the accuracy of pixel type decision was greatly improved.

4.2 FCN vs. Our Approach

The advantage of the FCN model is that the depth convolution can make good use of rich detail features of the image, which is obvious when extracting a target object with a large number of pixels. However, if the target object covers a large number of pixels even when one pixel contains several target objects, the use of depth convolution not only fails to achieve the purpose of extracting more detailed features but may also introduce more noise due to the expansion of the visual field, which affects the decision of pixel ownership. When using FCN to extract farmland and woodland from GF-2 images, although it occupies a large number of pixels in a farmland or woodland, the difference in pixel coverage due to a single plant is small and the advantage of the FCN cannot be played.

In contrast to the idea that the FCN model extends the field of view by deepening the hierarchy, the CENN model expands the field of view by increasing the "width". In the GF-2 image segmentation, using 3*3*3, 5*5 *3 and 7*7*3 three convolution kernels, the maximum observed area is about 49m2, which can cover most of the canopy. Therefore, the CENN model can not only fully exploit the features of the pixels themselves, but also fully exploit the spatial relationships between pixels, taking good account of the natural appearance of crops and trees continuously. In addition, the CENN model takes full account of the natural distribution of farmland and woodlands and is more advantageous in identifying pixels at corners.

From the experiment, we can conclude that the FCN model and the CENN model have similar segmentation accuracy when identifying the middle region of farmland and woodland. However, there are significantly more FCN model recognition errors when identifying the pixels in the corner regions, whereas the CENN model has almost no errors, which reflects the design advantage of the CENN model.

4.3 Deeplab vs. Our Approach

Compared with the FCN model, there are two significant improvements in the Deeplab model: (1) the deconvolution part is improved; (2) the network finally uses the Fully-Connected Conditional

Random Fields to refine the segmentation boundary. These two improvements are very beneficial in the segmentation of individuals with a large number of pixels. From the open literature, Deeplab's segmentation accuracy is better than that of the FCN model when identifying large objects such as buildings. The accuracy of the segmentation is due to the fact that Deeplab makes better use of the details of the image and the spatial correlation of the pixels over a larger area. However, when Deeplab was used to identify woodlands and farmlands, due to the pixel block details haven't changed much farmland, the information available is less, and the farmland and forest land within a wide range of spatial correlation is not strong, lead to Deeplab did not give full play to the effect of it. In the design, CENN fully considered the characteristics of single plant crops and small tree cover area, including fewer pixels and fewer details, these inclusions when viewed in reference with the fact that the plants usually appeared continuously in nature, ensured that the CENN model was applied in extracting the advantage of farmland and woodland.

5.Conclusions

This paper presents a CENN model that can extract farmlands and woodlands from GF-2 images. Compared with the classical DBN model, FCN model and Deeplab model, the CENET model we designed fully considered of characteristics of crops and trees in the GF-2 image. According to the characteristics of the model, the method of classification training is adopted in the model training so that the model can obtain sufficient discrimination ability and achieve the goal of extracting farmland and woodland with high accuracy from the high score 2 image. The paper also provides a method of using ROI for sample annotation, which can reduce the manual workload of marking and improve the marking efficiency.

The goal of CENN is to make up for the shortcomings of extracting models of farmland and woodland on the meter level images of GF1 and GF2 such as FCN model and Deeplab model. When the resolution of the image is better than the meter level, the CENN model no longer has an advantage because more plants cover a larger number of pixels and more detail appears on the image. At this time, it is no longer possible to use CENN for farmland and woodland extraction.

Acknowledgments: All The work was supported by the project of National Key R&D Program of China (2017YFA0603004)，the National Science Foundation of China (41471299，41671440)，Science Foundation of Shandong (ZR2017MD018，ZR2016DP01)，Open research project of Key Laboratory on meteorological disaster monitoring, early warning and risk management in characteristic agricultural areas of arid area(CAMF-201701).

Author Contributions: Chengming Zhang, Shuai Gao and Fan Yu Fan proposed and designed the technique roadmap, and performed the programming works; Chengming Zhang and Shujing Wan designed and performed the experiments; Qingdi Wei, Guang Wang, Qing Cheng and Dejuan Song collected the training and testing image data, and analyzed the experimental results.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. DONG Zhipeng, WANG Mi, LI Deren. A High Resolution Remote sensing Image Segmentation Method By Combining Superpixels with Minimun Spanning Tree[J]. Acta Geodaetica et Cartographica sinica, 2017,46(6):734-743.
2. Liu Dawei, Han Ling, Han Xiaoyong. High Spatial Resolution Remote Sensing Image Classification Based on Deep Learning[J]. Acta Optica Sinica,2016,36(4): 0428001(1)-0428001(9)
3. Liu C, Hong L, Chen J, et al.. Fusion of pixel-based and multi-scale region-based features for the classification of high-resolution remote sensing image[J]. Journal of Remote Sensing, 2015, 19(2): 228-239.
4. Jin Jing, Zou Zhengrong, Tao Chao. Compressed texton based high resolution remote sensing image classification[J]. Acta Geodaetica et Cartographica Sinica, 2014, 43(5): 493-499.
5. WU Zhaocong, HU Zhongwen, Zhang Qian, et al. On Combining Spectral, Textural and Shape Features for Remote Sensing Image Segmentation[J]. Acta Geodaetica et Cartographica sinica,2013,42(1):44-50
6. Miller, D.M.; Kaminsky, E.J.; Rana, S. Neural network classification of remote-sensing data. Comput. Geosci. 1995, 21, 377–386.

7. Mas, J.; Flores, J. The application of artificial neural networks to the analysis of remotely sensed data. Int. J.Remote Sens. 2008, 29, 617–663.
8. Camps-Valls, G.; Bruzzone, L. Kernel-based methods for hyperspectral image classification. IEEE Trans. Geosci. Remote Sens. 2005, 43, 1351–1362.
9. Mountrakis, G.; Im, J.; Ogole, C. Support vector machines in remote sensing: A review. ISPRS J. Photogramm.Remote Sens. 2011, 66, 247–259.
10. Pacifici F, Chini M, Emery W J. A neural network approach using multi-scale textural metrics from very high-resolution panchromatic imagery for urban land-use classification[J]. Remote Sensing of Environment, 2009, 113(6): 1276-1292.
11. Huang X, Zhang L. An SVM ensemble approach combining spectral, structural, and semantic features for the classification of highresolution remotely sensed imagery[J]. IEEE Transactions on Geoscience and Remote Sensing, 2013, 51(1): 257-272.
12. Liu C, Hong L, Chen J, et al.. Fusion of pixel-based and multi-scale region-based features for the classification of high-resolution remote sensing image[J]. Journal of Remote Sensing, 2015, 19(2): 228-239.
13. Yuan, Y.; Lin, J.; Wang, Q. Hyperspectral Image Classification via Multitask Joint Sparse Representation and Stepwise MRF Optimization. IEEE Trans. Cybern. 2016, 46, 2966–2977.
14. Wang, Q.; Lin, J.; Yuan, Y. Salient Band Selection for Hyperspectral Image Classification via Manifold Ranking. IEEE Trans. Neural Netw. Learn. Syst. 2016, 27, 1279–1289.
15. Bengio Y. Learning deep architectures for AI[J]. Foundations and Trends in Machine Learning, 2009, 2(1): 1-127.
16. Larochelle H, Bengio Y, Louradour J, et al.. Exploring strategies for training deep neural networks[J]. Journal of Machine Learning Research, 2009, 10(1): 1-40.
17. Jones N. The learning machines[J]. Nature, 2014, 505(7842): 146-148.
18. Nguyen, T.; Han, J.; Park, D.C. Satellite image classification using convolutional learning. In Proceedings of the AIP Conference, Albuquerque, NM, USA, 7–10 October 2013; pp. 2237–2240.
19. Wang, J.; Song, J.; Chen, M.; Yang, Z. Road network extraction: A neural-dynamic framework based on deep learning and a finite state machine. Int. J. Remote Sens. 2015, 36, 3144–3169.
20. Taormina, R.; Chau, K.W. Data-driven input variable selection for rainfall–runoff modeling using binary-coded particle swarm optimization and Extreme Learning Machines. J. Hydrol. 2015, 529, 1617–1632.
21. Liang, Z.; Shan, S.; Liu, X.; Wen, Y. Fuzzy prediction of AWJ turbulence characteristics by using typical multi-phase flow models. Eng. Appl. Comput. Fluid Mech. 2017, 11, 225–257.
22. Bellary, S.A.I.; Adhav, R.; Siddique, M.H.; Chon, B.H.; Kenyery, F.; Samad, A. Application of computational fluid dynamics and surrogate-coupled evolutionary computing to enhance centrifugal-pump performance.Eng. Appl. Comput. Fluid Mech. 2016, 10, 171–181.
23. Zhang, J.; Chau, K.W. Multilayer Ensemble Pruning via Novel Multi-sub-swarm Particle Swarm Optimization. J. Univ. Comput. Sci. 2009, 15, 840–858.
24. Wang,W.C.; Chau, K.W.; Xu, D.M.; Chen, X.Y. Improving forecasting accuracy of annual runoff time series using ARIMA based on EEMD decomposition. Water Resour. Manag. 2015, 29, 2655–2675.
25. Zhang, S.; Chau, K.W. Dimension reduction using semi-supervised locally linear embedding for plant leaf classification. Emerg. Intell. Comput. Technol. Appl. 2009, 948–955.
26. Wu, C.; Chau, K.; Fan, C. Prediction of rainfall time series using modular artificial neural networks coupled with data-preprocessing techniques. J. Hydrol. 2010, 389, 146–167.
27. Castelluccio, M.; Poggi, G.; Sansone, C.; Verdoliva, L. Land Use Classification in Remote Sensing Images by Convolutional Neural Networks. arXiv, 2015, arXiv:1508.00092.
28. Hu, F.; Xia, G.S.; Hu, J.; Zhang, L. Transferring deep convolutional neural networks for the scene classification of high-resolution remote sensing imagery. Remote Sens. 2015, 7, 14680–14707.
29. Wang, J.; Song, J.; Chen, M.; Yang, Z. Road network extraction: A neural-dynamic framework based on deep learning and a finite state machine. International Journal of Remote Sensing 2015, 36, 3144–3169.
30. Saito, S.; Yamashita, T.; Aoki, Y. Multiple object extraction from aerial imagery with convolutional neural networks. Electronic Imaging 2016, 2016, 1–9.
31. Noh, H.; Hong, S.; Han, B. Learning deconvolution network for semantic segmentation. Proceedings of the IEEE International Conference on Computer Vision, 2015, pp. 1520–1528.

32. Paisitkriangkrai, S.; Sherrah, J.; Janney, P.; Hengel, V.D. Effective semantic pixel labelling with convolutional networks and conditional random fields. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops, Boston, MA, USA, 7–12 June 2015; pp. 36–43.
33. Papandreou, G.; Kokkinos, I.; Savalle, P.A. Untangling local and global deformations in deep convolutional networks for image classification and sliding window detection. arXiv preprint 2014, arXiv:1412.0296.
34. Badrinarayanan, V.; Handa, A.; Cipolla, R. Segnet: A deep convolutional encoder-decoder architecture for robust semantic pixel-wise labelling. arXiv preprint arXiv:1505.07293 2015
35. Ioffe, S.; Szegedy, C. Batch normalization: Accelerating deep network training by reducing internal covariate shift. arXiv preprint arXiv:1502.03167 2015.
36. Liu, J.; Liu, B.; Lu, H. Detection guided deconvolutional network for hierarchical feature learning. Pattern Recognition 2015, 48, 2645–2655.
37. Volpi, M.; Ferrari, V. Semantic segmentation of urban scenes by learning local class interactions.Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops, 2015,pp. 1–9.
38. Simonyan, K.; Zisserman, A. Very deep convolutional networks for large-scale image recognition. ArXiv preprint arXiv:1409.1556 2014.
39. anboonyuen, T.; Vateekul, P.; Jitkajornwanich, K.; awawirojwong, S. An Enhanced Deep Convolutional Encoder-Decoder Network for Road Segmentation on Aerial Imagery. Recent Advances in Information and Communication Technology Series 2017, 566.
40. Kendall, A.; Badrinarayanan, V.; Cipolla, R. Bayesian segnet: Model uncertainty in deep convolutional encoder-decoder architectures for scene understanding. arXiv preprint arXiv:1511.02680 2015.
41. Chen, L.C., Papandreou, G., Kokkinos, I., Murphy, K., Yuille, A.L., Semantic Image Segmentation with Deep Convolutional Nets and Fully Connected Crfs, arXiv:1412.7062.
42. Long, J.; Shelhamer, E.; Darrell, T. Fully convolutional networks for semantic segmentation. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2015, pp. 3431–3440.
43. Gang Fu, Changjun Liu, Rong Zhou, et al..Classification for High Resolution Remote Sensing Imagery Using a Fully Convolutional Network. Remote Sens. 2017, 9, 498
44. Dolz, J., NeuroImage (2017), http://dx.doi.org/10.1016/j.neuroimage.2017.04.039
45. Vijay Badrinarayanan, Alex Kendall , and Roberto Cipolla, IEEE TRANSACTIONS ON PATTERN ANALYSIS AND MACHINE INTELLIGENCE, VOL. 39, NO. 12, DECEMBER 2017,2481-2495
46. Olaf Ronneberger, Philipp Fischer, and Thomas Brox. U-Net: Convolutional Networks for Biomedical Image Segmentation. arXiv:1505.04597v1 [cs.CV] 18 May 2015
47. Chen, L.C.; Papandreou, G.; Kokkinos, I.; Murphy, K.; Yuille, A.L. Deeplab: Semantic image segmentation with deep convolutional nets, atrous convolution, and fully connected crfs. arXiv preprint arXiv:1606.00915, 2016.
48. Haoning Lin, Zhenwei Shi*, and Zhengxia Zou, Maritime Semantic Labeling of Optical Remote Sensing Images with Multi-Scale Fully Convolutional Network, Remote Sens. 2017, 9, 480-501
49. Visin, F.; Ciccone, M.; Romero, A.; Kastner, K.; Cho, K.; Bengio, Y.; Matteucci, M.; Courville, A. Reseg: A recurrent neural network-based model for semantic segmentation. Proceedings of the IEEE Conference onComputer Vision and Pattern Recognition Workshops, 2016, pp. 41–48.
50. R Kingma, D.; Ba, J. Adam: A method for stochastic optimization. arXiv preprint arXiv:1412.6980 2014.